

## 歌声合成システム VOCALOID

剣持秀紀・大下隼人

ヤマハ株式会社 サウンドテクノロジー開発センター

hideki\_kenmochi@gmx.yamaha.com, hayato\_ohshita@gmx.yamaha.com

“VOCALOID”とは、ヤマハが開発した素片連結型の歌声合成技術およびその応用商品の総称である。すでにソフトウェア応用商品として数本が国内外で実際に市販されている。本稿では、VOCALOIDの基本構成、合成アルゴリズム、実際の商品でのユーザインタフェースを現時点での商品ラインナップを交えながら紹介し、最後に本技術の今後の見通しについても述べる。

### Singing synthesis system “VOCALOID”

Hideki Kenmochi, Hayato Ohshita

System for Advanced Sound Technologies, Yamaha Corporation

“VOCALOID” is a concatenative singing synthesis technology developed by Yamaha Corporation, and also a trademark for its application products. Its application software products are already commercially available in the world. In this paper, we would like to introduce its overview, its synthesis algorithm, and a graphical user interface used in its application software, mentioning its product line-ups. Finally we would like to its future prospects.

#### 1. はじめに

“VOCALOID”とはヤマハが開発した素片連結型の歌声合成技術およびその応用商品の総称である。2004年に最初のソフトウェア商品がリリースされて以来、数多くのユーザに利用されてきた。利用者数だけでいえば、現時点では世界で最も利用されている歌声合成ソフトウェアであると言っても差し支えないと思われる。

歌声を計算機上で人工的に合成しようとする試みは長年行われてきたが、実際の音楽制作に幅広く利用されるまでには至っていなかった。VOCALOIDでは、この点を踏まえ、音楽制作に「使える」歌声合成システムを目指して

開発を進めてきた。「使える」歌声合成システムのためには、得られる合成音声に以下の要件が必要である。(a)歌詞が聴取可能であること。(b)スムーズであること(接続時のノイズがないこと)。(c)自然であること(ブザー音的な合成音からの脱却)。また、合成音声そのものだけでなく、使いやすい楽曲制作環境も「使える」歌声合成システムのためには不可欠である。

現状ではVOCALOIDは残念ながらまだプロの音楽制作現場で日常的に使用されるまでには至っていないものの、利用範囲を限定すればアマチュアの愛好家が歌声を含む楽曲の制作のために簡単に利用できるツールとしての地位を確立しつつあると自負している。

VOCALOIDの基本構成を図1に示す。

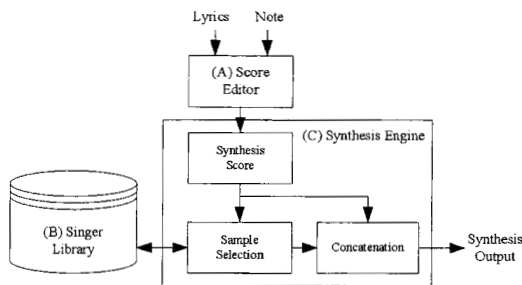


図 1 VOCALOID システム構成

それぞれの構成要素((A)スコアエディタ、(B)歌声ライブラリ、(C)合成エンジン)について、以下詳しく述べる。

なお、本稿で紹介するスコアエディタ、合成エンジンの構成は、いずれも本年(2007年)にリリースされた新バージョン(VOCALOID2)のものである。

## 2. スコアエディタ

スコアエディタは図 2 に示すような外観となっている。

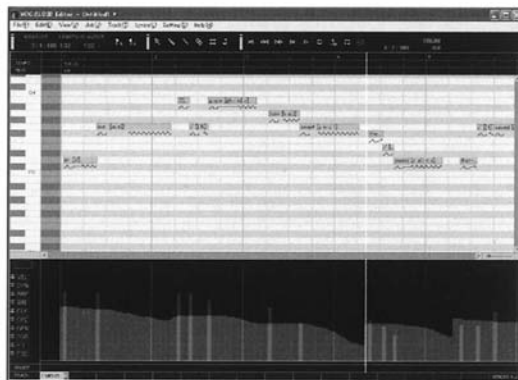


図 2 スコアエディタ

スコアエディタは音符、歌詞、表情を入力しやすいように設計されている。音符はピアノロールで入力する。

歌詞は音符をダブルクリックすることで入力可能となる。日本語の場合は仮名あるいはへボン式ローマ字で、英語の場合は単語そのものを入力する。歌詞は内部の発音辞書により、音

声記号に自動的に発音記号に変換される。発音記号を直接編集することも可能である。英語の場合は、複数の音節により構成される単語があるが、この場合は内部の発音辞書に含まれる音節区切りに従い、自動的に複数の音節に分割され、複数の音符に割り当てられる。

表情に関しては、音符に並んで表示されるアタックやビブラートの図形をマウス操作することでユーザーが好みの表情(歌い方)を音符単位で与えることが可能である。

スコアエディタ下の部分は、コントロールトラックとなっており、合成パラメータの細かい制御が可能となっている。

スコアエディタによる音符・歌詞の入力だけでなく、あらかじめ入力しておいた歌詞をもとにMIDIキーボードで演奏する機能も実装されている。これにより、キーボードで「歌う」という新しい音楽の演奏スタイルが可能となっている。

## 3. 歌声ライブラリ

歌声ライブラリは、実際の歌手の歌唱データを元にした音声素片を含むデータベースである。VOCALOID では音声素片の単位として、diphone と伸ばし音を使用している。歌声ライブラリには、対象とする言語で起こりうる全てのC-V、V-Cの組み合わせおよび全ての母音の伸ばし音が含まれている。

音声素片の元になるデータは複数のピッチで収録される。できるだけ多くのピッチで収録した方が合成音のクオリティが向上すると期待されるが、長時間歌い続けることによる声質の変化や歌手の疲れ(身体面および精神面)、歌手へのギャラなどを考慮し、ある程度のところでの妥協が必要となる。収録したデータは半自動で処理され、人手によるチェック、修正を経て歌声ライブラリが完成する。

#### 4. 合成アルゴリズム

合成エンジンは、歌詞、音符、表情その他のパラメータに従い、必要な音声素片を歌声データベースから取り出し、連結する。

歌声を合成する場合、素片を使用するタイミングが非常に重要になってくる。音節の先頭に子音が含まれる場合、Note ON のタイミングで子音を開始すると、タイミングの遅れとなって聞こえる。これを防ぐために、内部に「合成スコア」を持ち、音節の母音の開始部分が音符開始のタイミングに合うように素片の再生位置を調整し、タイミングがちょうど良く聞こえるようにする。図3に“Sing a song”([sIN a sO:N])という歌詞で歌う場合のタイミング調整の例を示す。

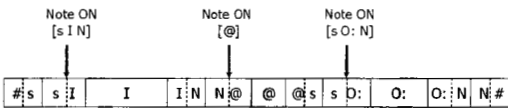


図3 素片タイミングの調整

合成スコアには、各時刻でのピッチや各種合成パラメータの変化も格納される。

素片を接続する時に、素片のピッチを所望のピッチに変換する必要があるが、たとえピッチを合わせたとしても、単純に接続しただけでは素片間の音色の違いから不自然な合成音やノイズが発生する。これを防ぐために、伸ばし音区間で、隣り合う diphone のスペクトル包絡を補間することで伸ばし音のスペクトル包絡とする。すなわち、“sing”([sIN])という歌詞の音符の伸ばし音部分のスペクトル包絡は、[s-I]の最終フレームと[I-N]の最初のフレームのスペクトル包絡を時間的に補間することで求められる。これにより原理的に接続部で音色の突然の変化が発生しないようになっている。diphone 区間は、素片のスペクトル包絡をそのまま使用する。

ピッチ変換はスペクトルを周波数軸上でス

ケーリングすることで行われる。倍音に相当するピークの近傍のスペクトルの形状はできるだけ元のものを保つようにスケーリングが行われる。また、ピッチを変換するために必要な位相の補償も行われる。

所望のスペクトル包絡に合うようにピークの強度が調整される。

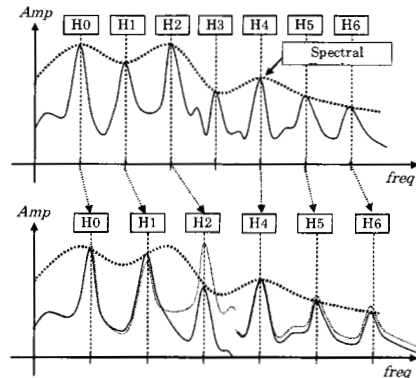


図4 ピッチ変換と音色の調整

#### 5. VOCALOID 応用商品

VOCALOID の応用ソフトウェア商品は、ヤマハからではなく、ヤマハとライセンス契約を結んだサードパーティーの商品として発売されている。Zero-G Limited (イギリス)[4]から2004年に、“Leon”(男声)、“Lola”(女声)、“Miriam”(女声)の3タイトルが発売され、クリプトン・フューチャー・メディア(株)(札幌市)[5]からは、2004年に“Meiko”(女声)が、2006年には“Kaito”(男声)が発売されている。これらのタイトルの違いは、パッケージに含まれる歌声ライブラリの違いである。

本年(2007年)には、VOCALOID の新バージョン (VOCALOID2) を搭載した商品が新たに発売された。PowerFX Systems AB(スウェーデン)[6]からは、“Sweet Ann”(女声)が既に発売されており、“Big-AL”(男声)も発売予定となっている。Zero-G Limited からは、クラシック系の

ソプラノ歌手を題材にした“Prima”(英語女声)が発売予定である。



© Zero-G Limited



© Crypton Future Media Inc.

### 図 5 VOCALOID 製品(Version1)

(上段右から Leon, Lola, Miriam,

下段右から Meiko, Kaito)

また、クリプトン・フューチャー・メディア(株)からは、実際に活躍する声優の声を題材にした「初音ミク」が発売されており、ネット上で話題を呼ぶとともに、音楽制作関連のソフトウェアとしては異例の売れ行きとなっている。



© Crypton Future Media Inc.

### 図 6 VOCALOID2「初音ミク」

## 6. 今後の見通し

VOCALOID の発売後、よく聞かれる反響に「やっぱり人間の声には勝てないね」というものがある。実際、現状では残念ながらプロの音

楽制作現場での定番として普及するには至っていない。つまり、「何時間もかけて合成音声を作りこむくらいなら、歌手を呼んできたほうが安いし早い」という論理には今のところは勝てないということである。この点についてどのように対応すべきか考える必要がある。

一方で、クリプトン・フューチャー・メディア(株)から発売された「初音ミク」は歌声自体にキャラクターを持たせることで、ヒットにつながった。この現象は詳しく分析する必要があると感じている。

また、ソロのボーカルではなく、[7]で筆者らが提案したコーラスの音声素片をもとにコーラスを合成するシステムであれば、「歌手を呼んできたほうが安いし早い」という論理に勝てる可能性がある。

## 7. 謝辞

VOCALOID の信号処理部分は、ヤマハと Pompeu Fabra 大学(バルセロナ)の Music Technology Group (MTG) との共同研究によって開発された。MTG のスタッフ、特に Jordi Bonada 氏、Alex Loscos 氏に感謝します。

## 8. 参考文献

- [1] Bonada, Loscos, Kenmochi, "Sample-based Singing-voice Synthesizer by Spectral Concatenation", Proc. of SMAC 03, 439-442, 2003.
- [2] Bonada et al., "Spectral Approach to the Modeling of the Singing Voice", Proc. of the 11th AES Convention, 2001.
- [3] Bonada et al., "Singing Voice Synthesis Combining Excitation plus Resonance and Sinusoidal plus Residual Models", Proc. of ICMC, 2001. Smith, J. O. and Abel, J. S., "Bark and ERB Bilinear Transforms", IEEE Trans. Speech and Audio Proc., 7(6):697-708, 1999.
- [4] <http://www.zero-g.co.uk/>
- [5] <http://www.crypton.co.jp/>
- [6] <http://www.powerfx.com>
- [7] 剣持, 大下, Bonada, Loscos, “コーラス音声の合成”日本音響学会講演論文集 1-Q-23 (2006-4)